PB82-156977

Nationwide Forestry Applications Program
Renewable Resources Inventory Project
Multiresource Inventory Methods Pilot Test
(Phase I): Final Report

Earth Satellite Corp.
Berkeley, CA

Prepared for

Forest Service
Houston, TX

15 Oct 80

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. NFAP-209	2.	3. Recipient's Accession No. PB82 156977
4. Title and Subtitle MULTIRESOURCE INVENTORY METHODS PILOT TEST (Phase I): Final Report			5. Report Date Date Published October 15, 1980
			6. 032045
7. Author(s) Earth Satellite Corporation			8. Performing Organization Rept. No.
9. Performing Organization Name and Address Earth Satellite Corporation Forestry Division 2150 Shattuck Avenue Berkeley, CA 94704			10. Project/Task/Work Unit No. AgRISTARS, RRI Project
			11. Contract(C) or Grant(G) No. (C) 53-3187-9-45 (G)
12. Sponsoring Organization Name and Address Nationwide Forestry Applications Program USDA, Forest Service 1050 Bay Area Blvd. Houston, TX 77058			13. Type of Report & Period Covered Final
			14.

15. Supplementary Notes

COLOR ILLUSTRATIONS REPRODUCED IN BLACK AND WHITE

16. Abstract (Limit: 200 words)

The Forest Service Multiresource Inventory Methods Pilot Test is a planned three phase effort, intended as a demonstration of Landsat satellite technology to supplement present methods of conducting recurrent inventories over large areas. The project creates an opportunity to apply earth resources satellite technology for testing and evaluating its potential for use by resource managers and planners. This Phase I Final Report contains a management summary of the results of Phase I of the Multiresource Inventory Methods Pilot Test. This report highlights a series of volumes that document Phase I. These volumes are:

- Volume I: Multiresource Inventory Design & Sampling Network
- Volume II: Multiresource Analysis & Information System (MAIS) Concept Development
- Volume III: Evaluation of MAIS Processing Components, Kershaw County, South Carolina Feasibility Test

17. Document Analysis a. Descriptors

- inventory
- geographic information system
- Landsat

b. Identifiers/Open-Ended Terms

- multiresource inventory methods
- Multiresource Analysis & Information System
- Landsat classification

c. COSATI Field/Group Forestry (Remote Sensing)

18. Availability Statement:	19. Security Class (This Report)	21. No. of Pages
Release Unlimited	Unclassified	41
	20. Security Class (This Page) Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse OPTIONAL FORM 272 (4-77) (Formerly NTIS-35) Department of Commerce

PB82-156977

AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

October 15, 1980

NATIONWIDE FORESTRY APPLICATIONS PROGRAM
Renewable Resources Inventory Project

Multiresource Inventory Methods Pilot Test (Phase I):

FINAL REPORT

EARTH SATELLITE CORPORATION
Forestry Division, Berkeley, California

USDA FOREST SERVICE
Houston, Texas

FOREWORD

This volume, together with Volumes One, Two, Three and Four, contains the results of the Multiresource Inventory Methods Pilot Test (Phase I). The principal authors of this set of volumes were Dr. Philip G. Langley, Dr. Jan van Roessel, Dr. Charles Sheffield, and Mr. Michael C. Place, of Earth Satellite Corporation.

Valuable assistance, discussions, and counsel were provided throughout this work by members of the U.S.D.A. Forest Service and associated consulting groups. We would especially like to acknowledge the contributions of Dr. Frederick P. Weber and Mr. Roger Pelletier of the Nationwide Forestry Applications Program, the work of the Program Management Staff within that group, and the work of personnel from the Renewable Resources Evaluation Units located at the Southeastern Forest and Range Experiment Station, Asheville, North Carolina, and the Intermountain Forest and Range Experiment Station, Ogden, Utah.

MULTIRESOURCE INVENTORY METHODS PILOT TEST

EXECUTIVE SUMMARY - PHASE I

This report summarizes the results of Phase I and Phase IA of the planned four-phase Multiresource Inventory Methods Pilot Test. The Pilot Test is an advanced demonstration of the use of Landsat satellite technology to supplement current methods of conducting recurrent inventories over large land areas, particularly those associated with multiresource estimation.

The intent of the Pilot Test is to determine the extent to which Landsat and associated geographic information system technologies developed over the past several years can facilitate, improve, or replace present methods in the face of increasing renewable resource information needs in the United States. An important requirement of the Pilot Test is that the technologies developed must be compatible with methods currently used for assessing renewable resources, particularly those used by the Renewable Resources Evaluation (RRE) Survey conducted nationwide by the USDA-Forest Service.

With the passage of the Resources Planning Act (RPA) of 1974, amended by the National Forest Management Act (FMA) of 1976, it became incumbent upon the RRE Survey (formerly Forest Survey) to broaden its assessment responsibilities on the Nation's forests and rangelands to include multiple resources, other than those related just to timber and related forest products. The RRE has responded by expanding their data collection activities to include those required for making multiresource assessments.

Corollary to the field data requirement is the need for a method to efficiently expand these data, in a timely way, into valid estimates applicable to state, county, national forest and other planning unit levels. Traditionally, these expansions have been accomplished by means of supplementary data gleaned from periodically acquired aerial photographs. In this application, the photographs provide the quantitative breakdown of total land area into its proportionate component uses.

In the southeastern United States, where Phase II will be conducted, the RRE field locations consist of a more or less permanent collection of sites that are periodically monitored for change and the acquisition of new kinds of resource assessment data. On the other hand, estimates of current land use areas, to which the field derived variables are applied, must be completely reconstituted for each RRE remeasurement cycle by means of new aerial photography. This involves the interpretation of thousands of new plots located on the latest set of readily available aerial photographs. Furthermore, the timing of each RRE reassessment is necessarily constrained by the frequency with which the photographs are taken by other agencies.

The RRE has responded to the increased responsibilities for multi-resource assessments set forth by the RPA and FMA, as far as its field data requirements are concerned, but is constrained by the periodicity of conventionally acquired aerial photography and the large human effort required for their interpretation.

Therefore, it seems clear that a Multiresource Analysis and Information System (MAIS) based on Landsat and related technologies must, at a minimum, provide the following:

1. Continuing accommodation of RRE field data
2. Relief from the present dependency on the long term periodicity of conventional aerial photography
3. More timely reassessments of changes in the resource base
4. A means to more effectively address environmentally related parameters such as sedimentation, disturbances and public use
5. A data management system which allows one to view the resource base in different ways for timely multiresource assessments at any time.

The concepts, rationale, and components required for the Pilot Test, as well as the results obtained from a trial run of data through the current MAIS, are detailed in four separate but interdependent volumes, titled:

1. Multiresource Inventory Design and Sampling Network
2. Multiresource Analysis and Information System (MAIS) Concept Development
3. Evaluation of Multiresource Analysis and Information System (MAIS) Processing Components, Kershaw County, South Carolina, Feasibility Test
4. Phase II Implementation Plan.

The contents of these four documents are summarized in Volume V, the Multiresource Inventory Methods Pilot Test - Phase I Final Report.

Results of Phase I and IA

A preliminary version of the MAIS developed during Phase I was assembled and tested in Phase IA using real data pertinent to Kershaw County, South Carolina. The details of these tests are included in

Volume III, the component evaluation document. The results from these tests, outlined below, appear to be most promising.

In summary, they indicate that:

1. Data derived from Landsat by means of computer methods, when combined with data from relatively few aerial photos, can be used to replace the human interpretation of thousands of air photo plots to estimate land use acreages.

2. When assisted by a computer oriented geographic information system, the relatively small amount of air photo data needed in (1) above can be effectively combined with other data sources, such as topographic and edaphic, to derive variables needed in certain multiresource assessments. These include sedimentation (as expressed by soil erosion potential), disturbances, and others associated with spatial dependencies such as public use, utilities or transportation infrastructures.

3. The total MAIS concept, including the Upper and Lower Level information systems as well as the Estimation Subsystem linking them, can provide most, if not all, of the capabilities for multiresource data collection, analyses, reporting and mapping as called for by the Program Management Staff of the Nationwide Forestry Applications Program.

The conclusions stated above emanate from the results of the Phase IA testing for component evaluation. The specific results obtained so far include the following:

1. It has been verified that the basic concepts of data flow through the MAIS are viable. That is, when data from various

sources are properly entered into the system, valid results are obtained.

2. Data acquired from Landsat, air photo, map, and field locations have been accurately registered in a common framework by means of practical, machine-assisted methods employing geographic information system technology.

3. The estimate of current annual increment for Kershaw County, South Carolina, and its associated sampling error obtained by the most recent RRE Survey have been duplicated independently by means of the current MAIS and existing RRE field data. This is a most significant result because, in the beginning of Phase I, few people believed it was possible to address this variable effectively by means of Landsat technology.

4. The total land area of Kershaw County has been significantly segregated into ten land use classes including three forest type classes (conifer, hardwood and mixed). This capability will provide the means to address at least three of the six variables required in the Pilot Test; water, land base change and grass (a component of winter range to be evaluated in Phase III).

5. The areas of the three forest type classes estimated in (4) above were further segregated into two tree size and two tree density classes with a high statistical significance level. This is important for estimating current annual increment and other timber related parameters.

6. The acreage estimate for the land use category "grass" was obtained to an extremely high level of significance. This

will be important for the assessment of winter range in Phase III.

7. The proportion of total land area assigned to "water" was nearly perfect with a correlation coefficient approaching unity between Landsat and air photo derived data.

8. The Lower Level GIS provides a totally new method for assessing soil erosion potential for use in addressing the sedimentation variable, one of the required six. High correlations were obtained between a sample of soil erosion potential maps and seven Landsat spectral classes. However, more observations to be acquired in Phase II, will be required to adequately assess the relative contribution of Landsat data.

In summary, the preliminary results obtained for Kershaw County, South Carolina support the design concepts for the MAIS as integrated with existing RRE field data. Estimates for certain classes of parameters, such as sedimentation (as a function of erosion potential) are within reach while the more conventional estimates of timber resources and land use can be made with the same or better precision when compared to current methods. The proposed techniques are both robust and flexible and, while high initial Landsat classification accuracies are desirable, they are not essential to obtain good results in the final analyses over a range of multiresource variables. Phase II will provide the additional data, resources and time to fully test the MAIS concept in South Carolina, and if successful, pave the way for the Phase III operational test in the State of Idaho.

BGR 110?5-15054 24 MAY 75 N34-32 W079-57 P017 R036 KERSHAW CO. SC CLASS MAP WITH COUNTY BOUNDARY

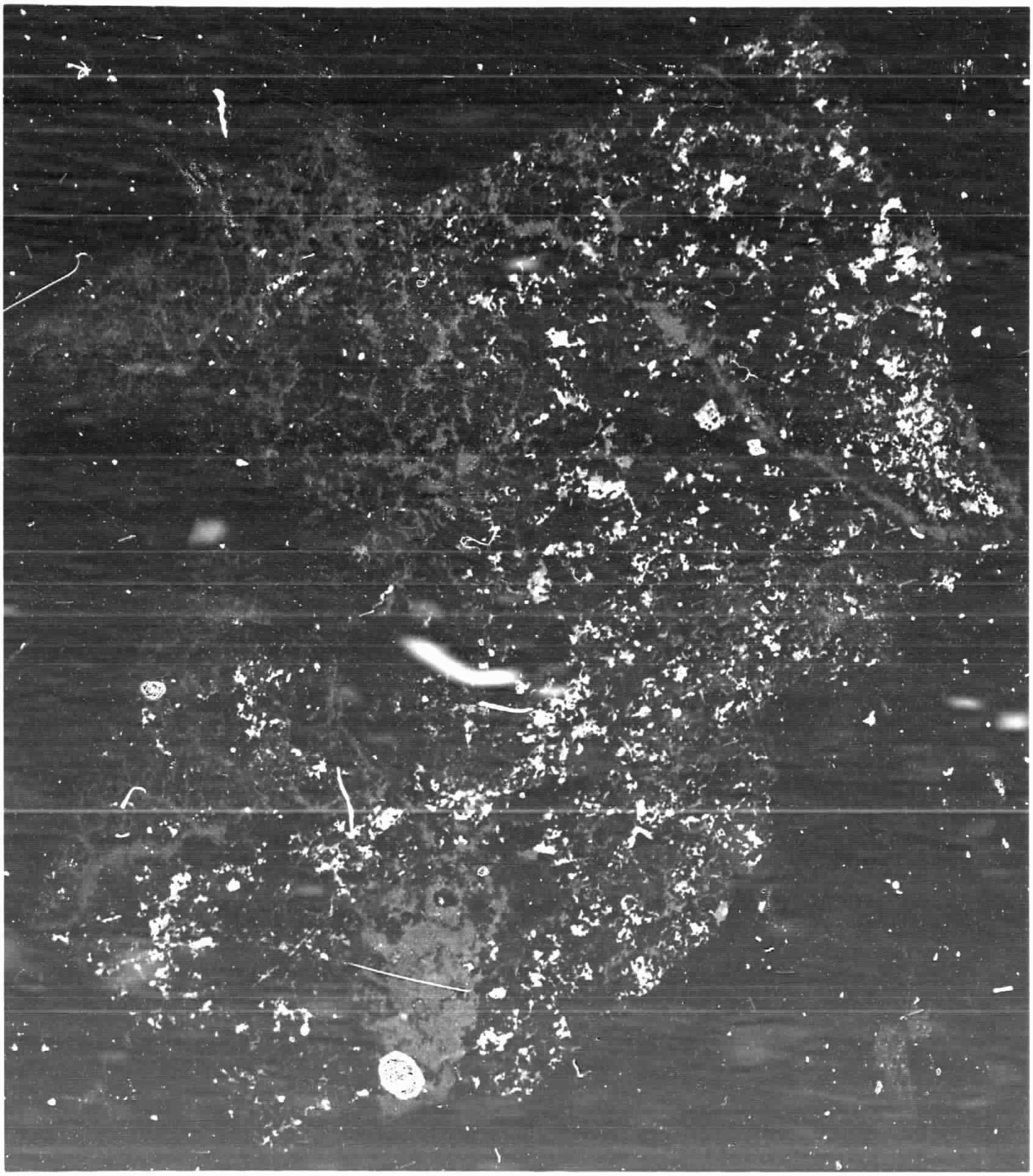

TABLE OF CONTENTS

1.0 INTRODUCTION

This volume constitutes Volume V (Final Report) of a series of volumes that document Phase I and Phase IA of a four phase Pilot Test for the application of Landsat data to problems of multiresource forest inventory.

The four phases of the Pilot Test were conceived as follows:

- Phase I: Development of conceptual framework for the Pilot Test, and evaluation of available methods and systems for potential use in the Pilot Test.
- Phase II: Application of preferred methods and systems to a multiresource operational test in a 16-county area of South Carolina.
- Phase III: Application of methods and systems to a multiresource operational test in an appropriate Western State, later designated as Idaho.
- Phase IV: Final evaluation and recommendations for the use of any resulting method or system in a general operational framework.

Phase I was initiated in the fall of 1979, with all the Pilot Test phases scheduled to occupy three years of development and application. During the performance of Phase I, the desirability emerged of a more detailed testing of concepts and methods in a practical test, before embarking on the large-scale production tests of Phases II and III. A Phase IA was therefore defined, in which methods and procedures evolved in Phase I were applied to real data for a single county (Kershaw County) of South Carolina.

Phase I and Phase IA are completed, and the results that have been obtained in them are documented in a series of Volumes, I through V, with the following contents:

Volume I:	Multiresource Inventory Design and Sampling Network
Volume II:	Multiresource Analysis and Information System (MAIS) Concept Development.
Volume III:	Evaluation of MAIS Processing Components, Kershaw County, South Carolina Feasibility Test.
Volume IV:	Implementation Plan for Phase II
Volume V:	Final Report

To these should be added a two-volume special purpose report developed for use in the Pilot Test (Key for Use in the Identification of Wildland Resource Features Through the Direct Visual Analysis of Landsat Multispectral Scanner Imagery). This report contains interpretive keys relevant to multiresource analyses made directly from Landsat, optical bar and conventional aerial imagery.

The volumes are separately bound, but they are mutually interdependent. The present volume has as its purpose three major functions. First, to set the context of the effort within the complete Pilot Test, something which will also be found in more detail in Volume I.

Second, to discuss the overall contents and relationships of Volumes I through IV, which contain the major technical discussion of the results of Phases I and IA; and finally, to point out the logic that led to the use in those volumes of particular approaches, out of the many possible approaches to Multiresource Analysis and Information System design.

It should be noted that although the present volume makes reference to Volumes I through IV, a full picture of project activities and project background can only be gained using all five volumes.

Volume I (June 20th, 1980) describes the requirements of the National Renewable Resources Evaluation Program, and discusses methods that have been employed by the U.S. Forest Service in performing resource inventories. It defines the major parameters that describe the multiresource aspects of Renewable Resources Evaluation (RRE) surveys, and discusses the sampling methods that have evolved for use in the RRE and other inventories. Methods of stratification, measurement and remeasurement are described in Section 2 of this volume.

Section 3 outlines the uses that have been made of remote sensing in forest inventories, covering both aerial photographs and space vehicle data. The use of remotely sensed data is described for derivation of expansion factors for samples, for provision of general information regarding the forest population, and for distributing estimates of totals to subareas of the population. Stratification, measurement variables, extrapolation, repartitioning, and mapping are all covered in this section.

In Section 4, the ideas of geographic information systems are introduced as tools for the processing of geographic and remotely sensed data. The discussion here is general, with specific details being deferred to Volume II. The questions of data combinations and data resolution are briefly discussed, again with more details left to Volumes II and III.

Section 5 discusses the logic for inventory design in the context of the Resources Evaluation Survey, the parameters to be estimated, and the data available from aerial photographs, field plots, and Landsat data. This section also explains the roles of Upper and Lower Level information systems in the proposed inventory process.

In Section 6, the actual sampling concept is defined. The logic for selection of primary, secondary, and ultimate sample units is given, and the stratification procedure is defined. Section 7 describes the variables which are to be estimated by the sampling procedures of Section 6, and Section 8 provides the mathematical basis for both estimation and mapping of discrete and continuous variables. Section 9 describes the use of the methods developed in both the multicounty/river basin and the national forest planning units.

Finally, Section 10 concentrates on the contribution of Landsat to estimation and mapping. Estimates of variance are developed from which the contribution of Landsat data to reduction of variance can be estimated, although these concepts were substantially expanded in Volume III.

Section 11 gives a list of relevant references.

<u>Volume II</u> (June 20th, 1980) concerns the hardware and software components required to conduct the investigations defined in Volume I. It provides details of both the Upper and the Lower level geographic information systems.

In Section 1, the main properties of the Upper Level (coarse resolution, image based) and the Lower Level (fine resolution, photo and map data based) geographic information systems are described, together with their main interfaces. Section 2 summarizes the computer hardware, photogrammetric hardware, and related facilities needed to support the information systems.

Section 3 gives the Upper Level geographic information system in full detail, describing first its general capabilities, its choice of coordinate systems, and the choice of a grid cell structure rather than a polygon structure. This is followed by a discussion of data base

concepts and the methods of file handling that should be applied. The concept of control units is introduced and discussed, followed by a brief description of file manipulation methods. Next, the major characteristics of the Landsat data classification are described, and the preprocessing that must be applied before Landsat can be used in conjunction with other data, is discussed in detail. This section includes an analysis of both radiometric and geometric corrections that must be applied to Landsat data before they can be used.

The next sub-section discusses the use of digital terrain data, and describes its format and its processing. The use of terrain data tapes to produce slopes and aspects is discussed here, followed by a summary of procedures that must be used to convert paper maps via digitization to a form that can be used with the MAIS. The discussion includes consideration of arc to cell conversion methods, of map unit extraction, and of methods to provide combined analyses of multiple layers of data. Included in the latter class of methods are supervised and unsupervised classification, and both techniques are discussed. This section concludes with a brief discussion of output report generation, and the need for both image format and statistical table outputs.

Section 4 gives details of the Lower Level geographic information system. It begins with an overview that describes the general functions the Lower Level system must perform, and describes the main types of spatial data (polygon, lineal, and point) that the system must accommodate. Following that overview, the choice of coordinate system for compatibility of Upper and Lower Level systems is discussed, and the choice of the UTM system is made. Similarly, the logic for selection of a polygon format system is provided, pointing out the features of the Lower Level data sources that dictate such a choice.

The next subsection discusses the data base and file handling methods for the Lower Level information system. Map units, data layers, and resource units are defined, followed by the definition of methods by which different data sources will be entered into the system. This includes discussion of aerial photo data and of map data. Appropriate digitizing methods are reviewed, and basic requirements for satisfactory digitizing during the Pilot Test are provided.

The major processing components of the Lower Level information system are then described, including arc-to-polygon conversion programs, data entry, data base storage and retrieval, displays, map production and labelling, combination of data layers, report generation, map zone generation, logical operations between and within data layers, and the definition of a suitable language for defining sequences of commands within the Lower Level system.

Section 5 defines the relation between the Upper and Lower Level systems, referring back to the linear models of Volume I and Volume III.

Section 6 provides the list of appropriate references.

Volume III (September 30, 1980) contains the detailed mathematical models and their application to a practical test for the particular geographic area of Kershaw County, South Carolina.

Section 1 discusses the building in Phase IA of a loosely-linked system of processing components to form a "prototype system" adequate for the performance of a substantial methods test using real data. Interaction of components, and the ways that the prototype system falls short of being well-finished or complete, are briefly discussed, together with factors that will need future attention to create the "mature" information system. The different systems used as separate elements of

the "prototype system" are described, together with the special processing procedures that were adopted in order to be able to make use of them in Phase IA activities.

Section 2 describes the development of estimators for use in the Pilot Test, and points out some of the trade-offs for different types of estimators and sampling methods. The data structure that is implied for the system is summarized, and the way in which Upper and Lower Level data are tied together in the linear model is defined. The important new development by which Landsat and aerial photo data are tied together through the correlation of class area proportions is described in detail, together with a full discussion of estimators and variances. A complete mathematical description is provided, for both continuous and discrete variables.

The next subsection describes the computer programs that were generated to apply the theory of class area proportion correlations, estimates of variances, and tests of statistical significance. Program tests against previous problems available in the literature are reported, along with a discussion of computer memory requirements implied by application of the selected methods to the data of the Pilot Test.

This is followed by the discussion of the Phase IA test itself (Section 2.2). The test area is briefly described, along with the sample selection procedures for the choice of 210 Forest Service field plots, and for the use of 60 one-mile square aerial photo samples. Photo quality is discussed, together with photo interpretation methods that were employed on the project. Data entry methods for the use of photo data are described, together with a discussion of the use of ground control points. The procedures used for the different types of variables (e.g., soil loss, cover type) are defined.

Following a discussion of the use of Landsat data, classification procedures, and proportion extraction methods, the main results of the Phase IA analysis are presented. Results are discussed in three separate sections, corresponding to different variables of interest, namely, land use, current annual increment, and soil erosion.

Finally, Section 3 summarizes the results of all the tests performed, and concludes that the conceptual design is workable and the results support the value of Landsat data for the multiresource estimation problem.

Volume IV (September 30, 1980) sets the results of Volumes I to III into the context of Phase II Pilot Test performance, and develops an Implementation Plan for this phase.

Section 1 relates the Implementation Plan to the work of the other volumes, and establishes the general objectives that the Implementation Plan sets out to satisfy. It emphasizes the use of proven systems and accepted statistical methods.

In Section 2, the design factors that dominated the Implementation Plan development are summarized, and significant events that occurred during Phase I performance and influenced plan development are described. The logic that led to the provision of PI keys in Phase I is given.

Section 3 details the actual tasks that must be performed during the Phase II implementation, and Section 4 sets these into chronological order and lists the task dependencies. The critical path for project performance is identified.

Section 5 sets forth the system capabilities that must exist in order that the plan for Phase II can be implemented. It lists system features by component, for both the Upper and the Lower Level information

systems. Similarly, in Section 6 the data needed for the implementation are reviewed.

Section 7 estimates the effort needed to perform the implementation of Phase II, based upon the processing experience that was gained in the performance of the single county test of Phase IA. The estimates given are for man-power only, and are provisional since the degree of government involvement in the effort is not yet known. Similarly, the costs of computer time cannot be estimated since the choice of information system for processing of the Phase II Pilot Test data has yet to be made.

Section 8 lists appropriate references that affected Implementation Plan design.

Finally, the two-volume set, "Key for use in the identification of wildland resource features through the direct visual analysis of Landsat multispectral scanner imagery", provides color and color infrared stereo examples and associated keys for the use of Landsat imagery. Volume I contains the terminology employed, a discussion of the photography used in making the keys, the general description and detailed PI keys for a California test area. Volume II contains the general description and PI keys for South Carolina (Kershaw County) and Idaho test areas.

2.0 THE CHOICE OF APPROACH

Volumes I and III define the approach that was determined to be most suitable for conduct of the Pilot Test, and they give the statistical and procedural implications of that approach. In this section we describe some of the alternatives that were considered before making the final selection of approach. This is done for two reasons; first, to demonstrate that there are several ways of approaching the overall problem addressed in the Pilot Test; and second, because reviewers of earlier drafts of these documents have made it clear that the choice among the alternatives is not an obvious one.

2.1 Choice of Sample Dimensions

The logic that suggested we should use aerial photographs to provide an intermediate level of area between the one acre field plots and the multi-million acre area covered by a single Landsat image was described in Volume I. Roughly speaking, the smallest element of resolution that can be analyzed on a Landsat image is one acre. The sampling units chosen on the aerial photographs have an area of 640 acres. This is about the smallest size of sample unit that permits us to define Landsat classes and still have confidence that the class area proportions determined in classification of Landsat images have meaning.

A more difficult decision must be made regarding the way in which aerial photograph sample units are to be correlated with corresponding areas on the Landsat image. Four different approaches have been considered. Each of them is outlined in the following section.

2.2 Alternative Approaches

2.2.1 First Approach

This is at first sight the most straightforward and appealing. Using the aerial photographs, class types are delineated on the sample units using manual photointerpretation methods. In parallel, using either manual photointerpretation or computer classification methods, the same class types are delineated on corresponding Landsat sample units. The Landsat areas and aerial photo areas are registered to each other, and the potential of the Landsat data is estimated by evaluating the errors of omission or commission, treating the aerial photo as the "correct" designation of class type. This allows a simple table to be constructed, giving, for each class type and each sample unit:

1. The number of pixels correctly classified as being of that class type,
2. the number included in that class that are actually of some other type, and
3. the number in that class that were defined as belonging to some other class.

There are two difficulties with this approach. First, the results are sensitive to the accuracy with which one can register Landsat and aerial photo sample units. Slight misregistration will introduce significant errors of both omission and commission. Second, there is an implicit assumption in such an approach that class types recognizable as distinct on Landsat images will correspond exactly to class types on aerial photographs (and then, by assumption, on the ground). However, even in cases where there is clean separation

of classes in the Landsat classification process, it is not necessarily true that those classes correspond directly to any ground-based or aerial photo-based class types.

2.2.2 Second Approach

Once it is recognized that class types on Landsat may not bear a one-to-one correspondence with class types on aerial photos and on the ground, a second approach suggests itself. Instead of classifying Landsat in terms of the class types required for the final ground-based categories, we can classify Landsat into classes that are denoted only by class numbers, 1, 2, 3..., without attempting to say that these are unique and named class types on the aerial photography. This is preferable conceptually, but it runs into difficulties when we consider the way in which the aerial photo classes and the Landsat classes are to be correlated with each other. We can certainly construct a series of correlation coefficients showing how each numbered Landsat class correlates with each aerial photo class type, but in order to do this we again require good registration of Landsat to aerial photos. Further, this approach, like the first one, may apply to class types but it cannot be applied when we are dealing with continuous variables such as current annual increment.

2.2.3 Third Approach

Continuous variables can be dealt with by treating the aerial photo delineations as training areas for Landsat classification. In this approach, we first register the Landsat and aerial photo

sample units. Next, the delineations for continuous variables on the aerial photos are used to determine regression coefficients in equations for which the Landsat gray levels serve as the independent variables. For example, suppose that we have variable values assigned on the aerial photos for a particular sample unit. Let us suppose that the aerial photos have been registered with Landsat, so that we can assign a numerical value o. the continuous variable for each Landsat pixel. Then this value is used as the <u>dependent</u> variable, V(i), and the Landsat gray levels, G(i,j), in the four spectral bands as the independent variables, to build a regression equation of the form:

$$V(i) = \sum_j G(i,j).A(j)$$

(j = Landsat band index, and the index i runs over all pixels in the sample unit. A(j) = regression coefficient).

We now solve by least squares for A(j).

This approach is again subject to the limitation on registering Landsat accurately to aerial photographs. It also presumes a linear relation between gray levels and continuous variables, which we have no <u>a priori</u> reason to assume.

2.2.4 <u>Fourth Approach</u>

If we accept that class types on Landsat may not have direct correspondence to the class types on aerial photography, and if we

also seek a method that is insensitive to exact registration of Landsat and aerial photo data sources, we can proceed via the correlation of areas. In this approach, the delineation into classes is performed for the aerial photograph sample units, and the area of each class type is found by mensuration. Next, the classification of corresponding Landsat sample units is performed, either by manual interpretation or by computer classification, with or without collateral data being used. The classes are not identified with particular ground-based or aerial photo-based class types, but the total class area proportions in the sample unit are measured (if manual methods were used) or computed (if computer classification was used). The class area proportions derived from the aerial photos are then correlated with the class area proportions derived from the Landsat image. Except at the boundaries of the sample units, this method is quite insensitive to the accuracy of registration of the two data sources. This was the approach used in Phase IA.

The obvious question that might be raised about such an approach is this: since the correlation of areas does not use exact, point-by-point matching of Landsat and aerial photo samples, has a significant amount of useful information been lost by dealing with areas alone? In particular, does the insensitivity to misregistration more than make up for any loss of exact geometric relationships of class types?

The results reported in Volume III suggest that if there is any loss of information by working with areas, it is not enough to spoil the results obtained. Very satisfactory correlations were found between Landsat and the aerial photography.

2.3 The Choice of Processing System

Volumes I and II describe the overall features and detailed processing elements of the MAIS needed to carry out the approach given in full in Volume III, and in brief outline at the end of the previous section. In this section the main factors that led to the adoption of a two-level processing philosophy are outlined.

Geographic information systems (GIS) usually adopt one of two overall structures. They are either cell-based information systems, or polygon-based information systems. In a cell-based system, the geographic area is divided up into a regular array of rectangular cells, usually all of the same area. With each cell are associated all the physical descriptors relevant to that parcel of land contained within the cell's area. Different layers of data can be combined first by making sure that their cell grids are compatible, and then by forming for each cell lists of descriptors that are some logical combination of the descriptors from each layer. Processing of multiple data layers within a cell-based GIS is, therefore, simple and fast. As a compensating disadvantage, it is very difficult to represent high-resolution features (such as road or stream patterns) within a cell-based system. Either a road is represented with a coarse resolution, or the number of cells needed in the system becomes astronomical. In any attempt to give a detailed representation of a land area (even an area as small as a few square miles) the cell-based system is inefficient.

In a polygon-based system, each data layer is represented and stored in the computer as a set of distinct polygons, each one having straight line segments as its sides. A polygon-based system

allows one to represent map data efficiently, since most map features such as contours, roads, streams, and class boundaries can be represented well using polygons. Combination of data layers, however, is much more complex with polygon-based systems. Before a combined layer can be created, it is necessary to form the set of polygons that is generated when all the line segments from both original polygon data layers are included. This is a logically complex problem, and even an efficient solution will require much computation for the combination of several data layers.

The question that had to be addressed during the MAIS design was, therefore, should the MAIS be built around a cell-based or a polygon-based structure?

This is not an easy decision. First, the Landsat data, in the form available on Computer Compatible Tapes, is essentially a cell-based system, with a cell size of 57 x 79 metres. Each Landsat image contains about 7.6 million such cells. However, even this high number of cells does not provide enough resolution to represent features such as roads and streams that must be recognized on the aerial photography. The latter data source is represented much better by a polygon-based system. We thus have two essential data sources, each best represented by a different type of information system.

The approach that was adopted in the MAIS recognizes the need to preserve the main characteristics of both types of data. The cell-based, low resolution Landsat data are handled by a cell-based processing logic, and constitute the "Higher Level" information system. The polygon-based, high resolution aerial photography is

handled by a polygon-based processing logic, and constitutes the "Lower Level" information system. The link between the two data sources is provided by the Estimation Subsystem, which, because it deals only with class area proportions, contains no explicit geometric structure, and is thus neither cell or polygon based.

In order for the two-level approach to work it is of course necessary to adopt a common map projection for both types of data. Landsat data are originally in a Space Oblique Mercator projection, and the aerial photos are in no standard projection. The chosen UTM coordinate system is one that is compatible with both, and also one that is in widespread general use.

Full details of the coordinate system, coordinate conversions, and photo rectification are provided in Volumes I, II and III.

3.0 APPLICATION OF THE METHOD

When Phase I of the Pilot Test was initiated in September 1979, it was envisioned that existing components would be selected for inclusion in a MAIS which, in turn, would drive conventional sampling strategies. In these applications, Landsat and collateral data would provide the basis for stratification before sample selection as well as post-sampling extrapolation of the results to various areas of the population.

As we progressed into the project, studied the state-of-the-art of Landsat applications in forestry and observed the modus operandi of the RRE in the Southeast, it became increasingly clear that we were dealing with a resource monitoring, statistical modeling and data management problem as well as a sampling program. Sometime after a visit to the RRE unit in Asheville, North Carolina in late January of this year, it became apparent that some kind of statistical prediction model linking the Upper and Lower Level geographic information systems would be needed to make the whole concept of the MAIS work. This model would translate Landsat derived data into information that was meaningful at the aerial photo and field levels.

However, it was not until Phase IA that the full significance of the "Linear Model" became forcefully clear. During the course of Phase IA work, it was discovered that the "Linear Model" was the most underrated of the three major subsystems, the Upper and Lower Level GIS's being the other two. Consequently the "Linear Model" was redesignated the "Estimation Subsystem" to more accurately reflect its status and complexities in the overall MAIS.

A major part of the effort in Phase IA was necessarily devoted to the development of a functioning Estimation Subsystem and produce a

prototype MAIS. To accomplish these objectives some new concepts in software design were required to make the Estimation Subsystem functional. This was necessary, however, to answer the broader question "Will the MAIS concept work as presently conceived?" Based on the results of Phase IA, we believe the answer is not only affirmative but a new plateau may have been reached in the application of Landsat and geographic information system technologies to renewable resource assessment problems.

3.1 The Prototype MAIS

The MAIS consists of three major subsystems:

1. The Upper Level GIS,
2. the Lower Level GIS and
3. the Estimation Subsystem.

In the Upper Level subsystem, the resource base is represented in its entirety in the form of Landsat classification images, possibly combined with collateral data describing soils, topography and other features at the same level of resolution. The data in the Upper Level system is frequently updated as additional Landsat coverages are obtained. The Lower Level GIS stores much more detailed information at a high resolution but only of selected sample areas. The data in the Lower Level Subsystem is also kept up-to-date but is mostly based on maps, aerial photographs and ground data.

The Estimation Subsystem ties together the data in the Upper and Lower Level Subsystems and produces the multiresource estimates. As the MAIS is presently applied, the emphasis on random samples is shifted from the Lower Level to the Upper Level where new samples can be drawn or complete enumerations of populations can be easily made. In this

configuration, the error distribution within the estimation model is relied on for appropriate random effects.

3.2 Kershaw County Test

3.2.1 Sample Selection

In Phase II, sample selection will consist of a random drawing of points, described in UTM coordinates, from throughout the 16-county test area. in South Carolina. In Phase IA, however, some tests on the accuracy of transferring coordinate locations through the system were required. The coordinate locations of 210 existing RRE field plots in Kershaw County, having already been determined, provided a basis for these tests, Therefore, sample selection for the Kershaw County mini-test began with these RRE plots. In the MAIS, these are termed Ultimate Sample Units (USU's). In addition, a random sample of 60 USU's was drawn from the 210 to provide the basis for defining Secondary Sample Units (SSU's). Around each USU drawn, a one-mile square SSU was randomly located such that the USU was contained somewhere within the SSU boundary when oriented N-S and E-W. In Phase II, the SSU's and USU's will be completely separate sets of sample locations.

3.2.2 Lower Level GIS

3.2.2.1 Photo and Map Interpretation of SSU's

For this step, the SSU locations were sketched on existing 1:20,000 scale aerial photo enlargements. The details for the precise registration of the SSU boundaries in the Lower Level GIS can be found in Volume III, Section 2.2.2.1.3.

In preparation for photointerpretation, a list of land use classes was prepared consistent with the requirements prescribed in the RFP for the Pilot Test. Then a trip was taken to Kershaw County by EarthSat personnel to inspect and evaluate a number of sites for PI training purposes. After a training period, each of the 60 one-mile SSU's was exhaustively partitioned into its land use and forest type components. Each delineation was then classified according to the established definitions for land use, forest type, stand size and density. Finally, the SSU data were digitized and entered into the Lower Level GIS (see Volume III, Figures 2 and 3, pp 47,48).

In preparation for estimating soil losses by means of the Universal Soil Loss Equation (USLE), a topography layer was digitized for a subset of 12 SSU's. The contour data for this activity were obtained from U.S.G.S. 7 1/2' quadrangle maps covering a part of Kershaw County. By means of the Lower Level GIS (e.g., EarthSat's LANDPAK system) slope classes were generated from the contour data for use in the USLE. The description of how the remaining data items needed for entering into the USLE may be found in Volume III, Sections 2.2.2.1.4 and 2.2.2.1.5.

3.2.3 Upper Level GIS

The upper level functions used in Phase IA included Landsat preprocessing, unsupervised classification, image registration, statistical tabulation and map generation.

3.2.3.1 Landsat Preprocessing

Landsat scene 11035-15054, May 1975, containing the data for Kershaw County was chosen for the test. Preprocessing was performed by means of EarthSat's program CCTRFM and consisted of three functions: reformatting, scan line suppression and geometric correction (Volume III, Sec. 2.2.2.2.1).

3.2.3.2 Classification

Classification of the preprocessed image was performed by means of the ISODATA algorithm. The classification terminated cleanly after generating 14 clusters over the entire scene. The proportional breakdown of the scene into the 14 clusters is given in Volume III, Table 4. These were later collapsed into ten classes and ultimately into 7 classes (Volume III, Table 5) for the final analyses.

3.2.3.3 Image Registration and Data Extraction

In preparation for entering sample data into the Estimation Subsystem, it was necessary to extract the classified Landsat data precisely in geographic registration with the Lower Level SSU data. At the Upper Level, these sample units are referred to as Primary Sample Units (PSU's). Since both the Lower and Upper Level GIS data are stored in UTM coordinates, the sample unit boundary descriptions are transferred from one level to the other with relative ease.

In Phase IA, the coordinates for the 60 sample units were transferred from the lower to the Upper Level system. Then, the classification data for each PSU was extracted and the proportions of each PSU falling in each of the seven and then ten Landsat data classes were computed (Volume III, Sec. 2.2.2.2.4). Finally, the boundary of Kershaw County was digitized and passed to the Upper Level System where the classification data for the entire county were extracted and tabulated also.

3.2.4 Estimation

For purposes of testing the validity of the MAIS in Phase 1A, three primary categories of variables were addressed by means of the prototype MAIS. These were land use, current annual increment and soil erosion potential.

3.2.4.1 Land Use

Six aggregations of land use class proportions were estimated. These aggregations are listed in Volume III, Table 6, p. 66. The proportion of Kershaw County falling into each land use class based on 7 and 10 Landsat spectral classes are given in Volume III, Table 7. The test statistics resulting from the land use estimates are:

TABLE 1

STATISTICS FOR LAND USE CLASSIFICATION

	7 Spectral Classes	10 Spectral Classes
R	0.7677	0.7740
F	11.68	8.29
Tabled $F_{.05}$	1.00	1.27
d.f.	163; 477	90; 450

3.2.4.2 Forest Type

Three additional combinations of forest type and land use breakdowns were estimated for comparative purposes (Vol. III, Tables 8,9,10). Forest Types 1 included two tree size classes, Forest Types 2 included two tree density classes and Forest Types 3 included the size and density class breakdowns simultaneously. The F statistics for these combinations using the seven Landsat spectral classes are:

TABLE 2

STATISTICS FOR FOREST TYPE CLASSIFICATION

	Forest Types 1	Forest Types 2	Forest Types 3
R	0.6813	0.7643	0.7635
F	7.14	11.49	11.71
Tabled $F_{.05}$	1.00	1.00	1.00
d.f.	91; 689	91; 689	91; 689

3.2.4.3 Grass

Because the evaluation of winter range will be of specific interest in Phase III, a special grass category was estimated by itself with all other land use classes lumped into one non-grass category. The results are as follows:

TABLE 3

STATISTICS FOR CLASSIFYING GRASS

	7 Spectral Classes	10 Spectral Classes
R	0.9842	0.9828
F	285.21	195.15
Tabled $F_{.05}$	2.20	2.04
d.f.	7; 53	10; 50

3.2.4.4 Water

The estimated proportion of total land area in water was near perfect with the following statistics:

TABLE 4

STATISTICS FOR CLASSIFYING WATER

	7 Spectral Classes	10 Spectral Classes
R	0.9994	0.9993
F	21,989	14,707
$F_{.05}$	2.20	2.04
d.f.	7; 53	10; 50

3.2.4.5 Current Annual Increment (CAI)

Several estimates concerning CAI were produced during the test and the results of the tests are detailed in Volume III, Tables 17-20. The most significant result is the comparison of the total CAI for Kershaw County as estimated by means of the prototype MAIS and by the RRE in Asheville as reported in Forest Statistics for the Northern Coastal Plain of South Carolina (Craver, 1978). The two results are:

TABLE 5

CURRENT ANNUAL INCREMENT

	Total CAI	Standard Error	S.E.%
MAIS	24,671,410	1,865,234	7.56
RRE	24,435,000	1,777,540	7.25

The contribution of Landsat to the above MAIS estimate was calculated at 7.5% relative gain in precision (Volume III, p. 82).

3.2.4.6 Soil Erosion Potential

A unique approach was taken to the estimation of soil erosion potential (for use in the later analyses of sedimentation). The estimate of soil loss is obtained by means of the Universal Soil Loss Equation. However, the input variables for the equation are generated for the SSU's by means of the Lower Level GIS (LANDPAK). The method involves entering component data in the LLGIS by layers including land use (covertype), slope class, roads, disturbances, etc. Then, by means of the combinatorial capabilities of the system, a new classification map is generated for each SSU. This map identifies the area of each SSU falling in each of a pre-defined number of soil erosion potential classes that result from the combination of the primary layers in the database. These values are aggregated to produce an estimated soil loss value for each SSU. These estimates are then combined with the Landsat classification data by means of the Estimation Subsystem to produce the county estimates. The statistics obtained for classifying soil loss are:

TABLE 6

STATISTICS FOR ASSESSING SOIL LOSS

R	0.9502
F	8.39
$F_{.05}$	1.78
d.f.	42; 30

The associated sampling error on total soil loss for the county was relatively high (66 percent). However, this should come down in Phase II when a larger sample of SSU's will be obtained.

3.3 Summary

A prototype system test was undertaken to assess the workability of the proposed MAIS design. The most important aspect in assembling a system from the set of components is the integration of the components into a workable entity. It was realized that for the MAIS, the vital link in this process is the "Estimation Subsystem", and hence the evaluation effort was directed at trying the proposed techniques for this subsystem in a set of preliminary tests for one county.

The results seem to support the notion that the basic scheme works very well. Estimates which heretofore were impossible to make using conventional methods can be made using GIS and Landsat technology (erosion potential). Conventional estimates can be made with the same accuracy as current methods, hopefully at reduced cost. The proposed techniques seem to be both robust and flexible. High Landsat classification accuracies are by no means required, and using the class transformation concept, one can produce a wide variety of estimates to satisfy many needs. Valuable insight was gained into a possible structure for a permanent estimation subsystem. An automatic file handling system along the lines of the transaction concept outlined in the concept development document is highly recommended. Also, an automated report generator must be included in a future estimation subsystem.

In conclusion, it seems that the major reservations concerning the proposed techniques have been eliminated. Minor problems remain due to time and resource constraints. It is anticipated that they will be addressed in the Phase II effort.

4.0 RECOMMENDATIONS FOR THE PHASE II IMPLEMENTATION

The practical exercise involved in the use of real data for Phase IA provided a number of lessons that should be borne in mind when considering the execution of Phase II of the Pilot Test. In most cases these are lessons that could not have been predicted on conceptual grounds alone, since they involve factors that emerge only with the manipulation of real photographs, Landsat tapes, and collateral data.

4.1 Data

4.1.1 Landsat Data

Landsat data were required in Phase IA for the classification of scenes, where classification accuracy is of central importance if useful results are to be obtained. The first attempts at classification in Phase IA were performed using Computer Compatible Tapes that had been processed in the Master Data Processor at Goddard Space Flight Center (P format tapes). Although the tapes provided were supposedly already de-striped to remove grey level variations caused by variable detector sensitivities, in practice the tapes that we received still contained substantial striping, enough to render useless the classification results. Further, the re-sampling of the scene done by the Master Data Processor made it impossible to remove the effects of Goddard Space Flight Center's de-striping attempts, and thus impossible to substitute a more effective de-striping. It was therefore necessary to obtain unprocessed (A format) tapes, to remove the attempted de-striping that had been performed, and then to apply a de-striping algorithm that did in fact remove the variations due to varying detector sensitivity. This consumed both time and labor.

Based on the experience gained in Phase IA, we suggest that Master Data Processor output tapes in their present form be considered unacceptable for use in Phase II classification work, and that A-format tapes be required as the Landsat data type to be used in the Phase II work.

4.1.2 Collateral Data

Problems were encountered during Phase IA in obtaining necessary collateral data in a timely fashion. If Phase II is to be accomplished within the time frame allocated in Volume IV, it is essential that collateral data be assured according to a schedule agreed to at the project outset between the government and the Phase II contractor.

4.1.3 Aerial Photography

As remarked in Volume III, the available photo data for the Kershaw County test area varied in quality between fair and poor. It is suggested that prior to the performance of Phase II, a general review should be performed of the quality of aerial photo data for all 16 counties of the Pilot Test. Data deficiencies should be noted and this information made available at the beginning of Phase II implementation to the appropriate contractor personnel.

4.2 Computer Processing

The actual volumes of data and data processing called for in the Phase IA test was larger than originally anticipated. For Kershaw County, systems of 300 linear equations had to be solved, and matrices

of dimension 300 x 300 were inverted. The generation and processing of very large arrays is much more difficult with a fixed memory (as opposed to a virtual memory) machine, and would call for a good deal more programming effort to segment arrays and programs to use two-level storage.

For this reason, it is unwise to attempt to install the MAIS on a "small" computer system. In this case, minimal requirements for the computer that will support the MAIS are as follows (see also the appropriate section of Volume II):

- A directly addressable high-speed memory of at least 512,000 bytes, with twice this memory size desirable though not essential.
- At least 100 million bytes of disk storage.
- One or more tape drives.
- A complete FORTRAN language compiler, that satisfies ANSI standards.
- An operating system that supports multiple users and provides a virtual memory capability.
- Floating point hardware for arithmetic computations.
- Double precision arithmetic (i.e., 64-bit accuracy arithmetic).
- A long word length, of 32 bits or more. (This is satisfied by all large main-frame computer manufacturers, such as IBM (32 bit word), UNIVAC (36 bits), CDC (48 or 60 bits), and by a lesser number of minicomputers, such as the VAX 11-780 and PRIME 250, 550 and 750).

The absence of one or more of these features will complicate the provision of an MAIS that can handle the processing and computational load imposed by a full-scale 16-county Pilot Test. In the worst case, processing will become infeasible because of memory constraints, direct

addressing limitations, or computational accuracy. It should be noted that the inversions of large matrices encountered in running the Estimation Subsystem makes double precision arithmetic during that computational phase almost mandatory.